Deep Shapely Portraits

Qinjie Xiao
Zhejiang University, ZJU-Tencent Game and Intelligent Graphics Innovation Technology Joint Lab
qinjie_xiao@zju.edu.cn

Xiangjun Tang
Zhejiang University, ZJU-Tencent Game and Intelligent Graphics Innovation Technology Joint Lab
fcsx1tf@163.com

You Wu
Zhejiang University, ZJU-Tencent Game and Intelligent Graphics Innovation Technology Joint Lab
21860409@zju.edu.cn

Leyang Jin
The Chinese University of Hong Kong, Shenzhen
117010110@link.cuhk.edu.cn

Yong-Liang Yang
University of Bath
y.yang@cs.bath.ac.uk

Xiaogang Jin*
Zhejiang University, ZJU-Tencent Game and Intelligent Graphics Innovation Technology Joint Lab
jin@cad.zju.edu.cn

Figure 1: Given input portrait images with varying poses and expressions (bottom), our approach can automatically generate shapely portraits (top) that are better proportioned, by estimating the best reshaping parameter setting (called shapely degree) using deep learning.

ABSTRACT

We present deep shapely portraits, a novel method based on deep learning, to automatically reshape an input portrait to be better proportioned and more shapely while keeping personal facial characteristics. Different from existing methods that may suffer from irrational face artifacts when dealing with portraits with large pose variations or reshaping adjustments, we utilize dense 3D face information and constraints instead of sparse facial landmarks based on 3D morphable models, resulting in better reshaped faces lying in rational face space. To this end, we first estimate the best shapely degree for the input portrait using a convolutional neural network (CNN) trained on our newly developed ShapeFaceNet dataset. Then the best shapely degree is used as the control parameter to reshape the 3D face reconstructed from the input portrait image. After that, we render the reshaped 3D face back to 2D and generate a seamless portrait image using a fast image warping optimization. Our work can deal with pose and expression free (PE-Free) portrait images and generate plausible shapely faces without noticeable artifacts, which cannot be achieved by prior work. We validate the effectiveness, efficiency, and robustness of the proposed method by extensive experiments and user studies.

*Corresponding author.

CCS CONCEPTS

• **Computing methodologies → Image processing**.

KEYWORDS

portrait editing, shapely level, deep learning, face reshaping, datasets

1 INTRODUCTION

The human face is essential in representing personal identity and making a first impression. Most people wish that they look gorgeous in all their photos on Facebook or Instagram. As a result, portrait editing plays a significant role in many applications, such as social media, advertisement, visual effects, fitness incentives, etc. In practice, the facial attractiveness of a portrait image can be affected by two main factors, face texture and face shape [24, 27, 29]. The former is solely determined by the colors on the face and can be improved by direct color adjustments such as brightness enhancement [28, 36]. However, the latter cannot be simply resolved in this way, as the face shape can be varied according to one's body condition such as weight [43] in particular, even when the pose and expression of the face are fixed. Therefore, how to reshape face in the portrait image to improve facial attractiveness, while keeping the personal characteristics and the realism of the original portrait image, is an intriguing yet challenging problem.

Although it is straightforward to reshape objects in 2D based on image warping, the lack of underlying 3D information would easily cause artifacts due to unnatural deformation, especially for very familiar objects, such as human bodies and faces. This inspires researchers to leverage 3D deformation for realistic object reshaping, for instance, on human bodies [44]. In terms of face reshaping, the state-of-the-art method [43] utilize an irrational regression model (controlled by the differences of the body mass index (δBMI)) defined on a sparse set of facial landmarks to manipulate 3D face. However, this method may cause noticeable artifacts when performing on portraits that require large face shape adjustments and have large pose or expression variations, etc. The reason is that it relies on sparse feature points to reconstruct and deform a 3D face, and the resulting face may not lie in rational face space. Moreover, it is not automatic. Thus, the user is required to tweak parameters to achieve a satisfactory result manually. An intuitive approach is to represent 3D face using a parametric representation as 3D morphable models (3DMM) [2], then automatically estimate the face roundness (an indication of how shapely a face is), and finally make reasonable adjustment of the face shape according to that roundness estimate. Nevertheless, due to the representation limitation of 3DMM, it is challenging to preserve facial identity while making shape adjustment. Moreover, automatic roundness estimation requires labeled data across people with different gender, age, etc., which is currently not available. And more importantly, to ensure accurate and consistent annotations, it is preferable to label data from the same identity with different roundness values. But collecting such data is simply not feasible in practice, given the fact that the face roundness of an individual usually remains constant for an extended period, and even it changes, the rate is slow.

To address the above challenges, we develop an automatic shapely portrait editing method based on deep learning, achieving plausible reshaping results without changing the identity of the input portrait (see Figure 1). Different from Zhao et al. [43], our reshaping model leverages the parametric face representation based on 3DMM along with the state-of-the-art facial expression model. It enables dense and rational control of the 3D face by employing a reshaping optimization with identity and face roundness constraints. As a result, our model can simply use a single parameter δBMI to reshape 3D face in rational face space, enabling automatic shapely portrait generation for the first time while avoiding unexpected artifacts caused by previous work. We also present ShapeFaceNet, a portrait image dataset from 3,400 individuals. Each individual's portrait is adjusted with different face roundness using our reshaping method. The faces with most shapely or attractive degrees are annotated by human raters. Based on the dataset, we train a CNN-based estimator that can compute the best shapely level (face roundness), which can be used to reshape the input portrait. Moreover, we employ a resolution-free method inspired by [35] to warp the input image according to the reshaped 3D face model and its fitness with the image background, which avoids performance problem of Zhao et al. [43] as their warping algorithm is resolution-dependent.

We evaluate our method by extensive experiments on various portrait images and comparisons with the state-of-the-art. We also conduct a pilot study to validate that the shapely degree of the resultant portrait accords with human aesthetics. The results demonstrate the effectiveness, efficiency, and robustness of the proposed method. Overall our work makes the following contributions:

- We present a novel deep shapely portrait editing method that is simple, efficient, and stable. It can handle PE-free portrait images and automatically generate rational, realistic, and shapely faces to be better proportioned without noticeable artifacts while preserving personal identity.
- We present a reshaping model based on an optimization method that leverages dense and rational constraints concerning 3DMM, face identity, and face roundness, achieving plausible reshaping results.
- We create the first shapely portrait dataset ShapeFaceNet, which comprises portrait images from 3,400 individuals and have different shapely degrees annotated according to public aesthetic criteria. Based on this, a CNN-based shapely level estimator is proposed to estimate the shapely degree for automatic shapely portrait editing.

2 RELATED WORK

Portrait retouching, which has wide applications on social media and entertainment, is highly anticipated by academic and business sectors. Commercial software such as Adobe Photoshop is widely used but requires professional skills. To facilitate the portrait retouching process, many works have been proposed to edit the face texture, shape, or both. *Portrait texture editing methods* such as style transfer [36] transforms the style of a portrait to another. Same intuitive face properties like make ups [28] and lighting [37] can migrate as well. More lighting control of portrait images can be obtained by relighting under any given environment map [38]. Although the results of these methods are impressive, the portrait shape remains unchanged, limiting the scope of applicability. On the other hand, many *portrait shape editing methods* are proposed to further reshape input portraits. Shih et al. [35] reduce wide-angle distortions without affecting other parts of the input portrait image instead of reshaping the face itself. Interesting face-related applications such as personalized and photorealistic face caricaturing [16], caricature synthesis [25], and portrait characterization analysis and synthesis [34] are also presented. To improve the attractiveness of face shape, aesthetic-aware face reshaping method [24] edits the proportion of a frontal portrait image through a specific set of facial landmarks based on public aesthetic criteria. However, it is only performed in 2D and relies on accurate facial landmarks to adjust the placement of eyes, nose, etc., while we only change the face roundness in 3D, thus can preserve intrinsic face features/identity and handle large pose and expression variations. Liao et al. [27] extend the idea to 3D by employing empirical attractiveness criteria for 3D faces. It focuses on salient structure (facial features and contour proportions) adjustments instead of considering the underlying physiological structure of the human face, thus yielding artifacts. *Portrait texture/shape editing methods* are mostly based on generative models [21, 29]. They can change both the face texture and shape of portrait images. However, these methods are difficult to control as the reshaping parameters are not physiologically meaningful. The most relevant work to ours is [43], which presents a more general parametric reshaping method based on a soft tissue thickness regression model using sparse feature landmarks on the facial regions. It indirectly reshapes the input portrait image using the reshaped 3D face. However, this method is not automatic and may produce noticeable artifacts due to the irrational reshaped 3D face when dealing with large poses, expressions, and adjustments. To solve this problem, our method utilizes the underlying face roundness component of a 3DMM to represent the reshaped face, thus guarantees the rationality of the resulting face. We can also automatically estimate the shapely degree of a given portrait image and use it to guide the reshaping process.

CNN-based facial attractiveness learning. Deep learning makes it possible to learn high-level aesthetics-related facial features, and these features can be further applied for facial attractiveness inference. Various CNN-based approaches are proposed for the facial attractiveness prediction (FBP) task. Xie et al. [41] develop a face dataset with attractiveness ratings (SCUT-FBP) for automatic facial beauty perception. This benchmark dataset intrigues further deep learning based prediction methods, such as the psychology-inspired CNN-based method (PI-CNN) [42], self-taught learning [14], label distribution learning (LDL) [13], feature combination [9], and multi-task learning [15]. Rothe et al. [33] rate a specific person from a given portrait image based on the dataset on *howhot.io*. However, these methods are mainly designed to infer general attractiveness, which can be affected by factors such as gender, age, appearance, etc. Moreover, the dataset they used does not record the face shape changes of a specific person. Thus, it is not suitable to measure how shapely a portrait image is. Different from the above, we create a specific dataset, called ShapeFaceNet, from 3,400 individuals. It contains shapely degree annotations from human raters to meet the requirement for automatic shapely portrait generation.

Monocular face reconstruction. Monocular 3D face reconstruction aims to reconstruct 3D face geometry and texture from portrait image or video of a single view. It is well known to be ill-posed as shared by all single view reconstruction problems in 3D vision. Fortunately, such a problem can be constrained using the semantic information of 3D face provided by 3DMM, such as [2, 3, 19]. By taking such information as a prior, the underlying 3D face model can be faithfully estimated. Later, 3DMM is extended to include also 3D facial expressions [5, 26] for face editing applications. So far various 3DMM-based methods have been developed to reconstruct 3D face from single image [40] or video [23, 39]. The reconstructed 3D face can be enhanced by adding more elements such as hair style [6, 7], eyes [1], and fine-scale details [4, 32]. Besides, secondary components (eyes, teeth, tongue, and gums) can be included in the reconstructed face model [18] for gaming and VR applications. Unlike these applications, the edited face of our approach keeps its original pose and expression. As a result, our pipeline is less sensitive to the quality and accuracy of the reconstructed face geometry and texture. Hence a fast, simple, and robust monocular face reconstruction approach [19] is preferred.

Content-aware image warping. Image warping has been widely applied in various image editing applications, such as image stitching and reshaping [8, 10]. By embedding the input image into a content-aware 2D triangular mesh [43, 44], important contents can be warped with minimal visual distortion while retaining surrounding context. However, the performance of these methods is highly dependent on image resolution, which may prevent their usage in interactive applications for large images. Shih et al. [35] use a sparse constant control mesh to keep the background minimally distorted, resulting in high efficiency and resolution free performance. Inspired by their work, we present an efficient warping optimization with additional face boundary constraints to seamlessly overlay the reshaped face to the background.

3 OVERVIEW

In this work, we propose a novel method that automatically reshapes an input portrait image, making the resultant portrait shapely and realistic. Figure 2 gives an overview of the algorithmic pipeline. Given an input portrait image, we first detect the face regions and facial landmarks, and then utilize a monocular face reconstruction approach to reconstruct a 3D parametric face based on 3DMM (Section 4). After that, a deep-learning-based shapely level estimator trained on our new dataset ShapeFaceNet, is employed to estimate the shapely level of the input portrait (Section 6). Guided by the estimated shapely degree, the proposed reshaping model can drive a specific set of facial landmarks on the neutral shape (identity) of the reconstructed 3D face. We perform an optimization to reshape the neutral 3D face, and then combine it with the facial expression to generate the final reshaped 3D face (Section 5). Finally, a content-aware warping optimization is employed to seamlessly fuse the 2D face regions projected by the reshaped 3D face into the background, resulting in a shapely portrait with high realism (Section 7).

4 MONOCULAR FACE RECONSTRUCTION

We employ the face reconstruction approach by Huber et al. [19] to fit a parametric face model to a given portrait image. It is based on a meaningful face representation with a parameter vector $\boldsymbol{\mathcal{P}}$, which consists of head pose parameters (rotation $\mathbf{R} \in \mathbb{R}^{3\times3}$ and translation $\boldsymbol{t} \in \mathbb{R}^3$), facial identity coefficients ($\boldsymbol{\alpha} \in \mathbb{R}^{N_{id}}$, N_{id} = 63), and facial expression coefficients ($\boldsymbol{\beta} \in \mathbb{R}^{N_{exp}}$, N_{exp} = 6). Besides, we obtain the texture maps by orthogonally projecting the 3D face onto the image plane rather than estimating the diffuse skin reflectance, which dramatically reduces the reconstruction complexity while meeting our requirement for portrait reshaping. For an input portrait image, 81 parameters are reconstructed by the monocular face reconstruction. Next, we will present the parametric face representation and monocular face reconstruction in detail.

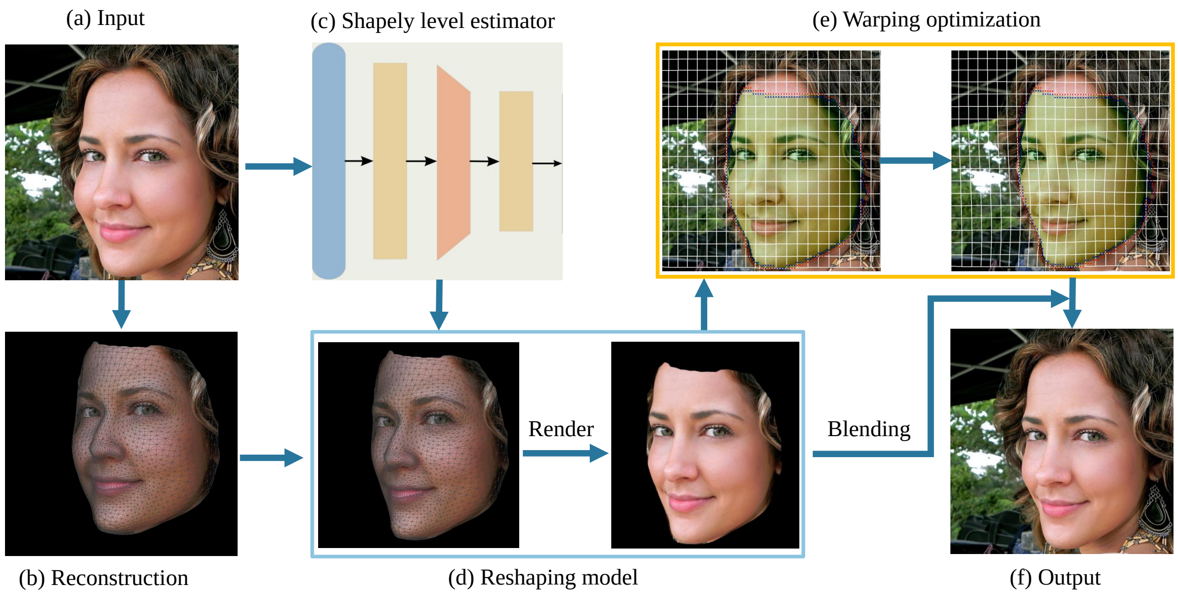


Figure 2: The pipeline of our deep shapely portrait method. Given an input portrait image (a), we first perform monocular face reconstruction to create a parametric 3D face (b). By estimating the shapely degree of the input face using a trained deep learning model on ShapeFaceNet dataset (c), we optimize the parametric 3D face (b) to obtain the shapely 3D face model (d)(left). After seamlessly fusing the rendered 2D face regions (d)(right) with the background using a fast, resolution-independent warping optimization algorithm (e), we generate the shapely portrait while keeping personal facial characteristics (f).

Parametric Face Representation. We represent the space of facial identity based on 3DMM [2], and the space of facial expression via the affine model of FaceWarehouse [5]. More specifically, a 3D human face $\boldsymbol{X}$ is formed by a mesh with n vertices, i.e., $\boldsymbol{X} = [\boldsymbol{x}_1; \boldsymbol{x}_2; ...; \boldsymbol{x}_n] \in \mathbb{R}^{3n}, \boldsymbol{x}_i \in \mathbb{R}^3$. The shape variation of human face is formulated as the summation of the mean shape $\bar{X} = [\bar{x}_1; \bar{x}_2; ...; \bar{x}_n] \in \mathbb{R}^{3n}, \bar{x}_i \in \mathbb{R}^3$, and the linear combination of a set of 3D face shape bases, which encode the per-vertex bias of the underlying template mesh as follows:

$$\boldsymbol{X}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \bar{X} + \mathbf{P}_{id} \cdot \boldsymbol{\alpha} + \mathbf{P}_{exp} \cdot \boldsymbol{\beta}, \tag{1}$$

where $\mathbf{P}_{id} = [P_1; P_2; ...; P_{N_{id}}] \in \mathbb{R}^{3n \times N_{id}}$ is the matrix consists of N_{id} identity shape bases computed by principle component analysis in Basel Face Model [2]. Similarly, $\mathbf{P}_{exp} \in \mathbb{R}^{3n \times N_{exp}}$ is the matrix composed of N_{exp} expression shape bases in FaceWarehouse [5].

Face Reconstruction. We employ a simple face reconstruction method to efficiently estimate the parameters $\mathcal{P}$ for the input portrait. The reconstruction is formulated as an optimization that combines terms for face alignment and statistical regularization:

$$E(\mathcal{P}) = w_f E_f(\mathcal{P}) + w_r E_r(\mathcal{P}), \tag{2}$$

where $\mathcal{P} = \{\mathrm{R}, \boldsymbol{t}, \boldsymbol{\alpha}, \boldsymbol{\beta}\}$. This enables the simple and robust reconstruction of geometry identity, facial expression, and head pose. We use 68 facial landmarks automatically detected by the implementation of [22] to define the sparse alignment term E_f when fitting a 3D face model to a 2D portrait image. The regularizer E_r enforces statistically plausible parameter values based on the assumption of normally distributed facial data. We refer to [23] for more details on the energy formulation.

5 RESHAPING MODEL

After monocular face reconstruction, we obtain the reconstructed 3D face $\boldsymbol{X}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$ based on identity coefficient $\boldsymbol{\alpha}_0$ and expression coefficient $\boldsymbol{\beta}_0$. We denote the identity component of face $\boldsymbol{X}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$ as $\boldsymbol{X}(\boldsymbol{\alpha}_0)$ by fixing expression coefficient $\boldsymbol{\beta}_0$. Then we have the following representation for each vertex in $\boldsymbol{X}(\boldsymbol{\alpha}_0)$:

$$\boldsymbol{x}_i(\boldsymbol{\alpha}_0) = \bar{x}_i + \mathbf{P}^i_{id} \cdot \boldsymbol{\alpha}_0, \tag{3}$$

where $\mathbf{P}^i_{id} \in \mathbb{R}^{3 \times N_{id}}$ is a sub-matrix of $\mathbf{P}_{id}$ corresponding to the x, y, and z coordinates of $\boldsymbol{x}_i$. We aim to generate a reshaped 3D

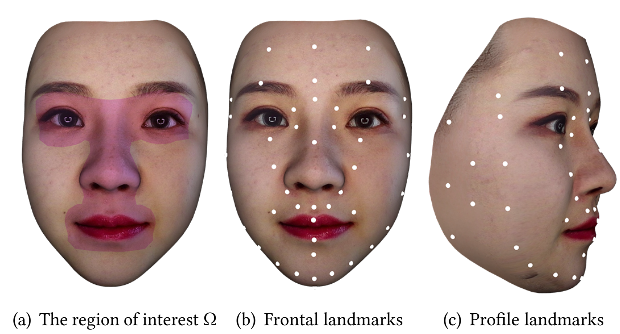


Figure 3: The red feature region (a) is used for face identity preservation. The facial landmarks illustrated in (b) and (c) are used for the regression model of soft tissue thicknesses. ©Hanqi Lv

face $Y = [\boldsymbol{y}_1; \boldsymbol{y}_2; ...; \boldsymbol{y}_n] \in \mathbb{R}^{3n}, \boldsymbol{y}_i \in \mathbb{R}^3$. To generate a satisfactory reshaped face, we need to preserve the identity and rationality of the input face when performing the reshaping operation. We first introduce the sparse model of Zhao et al. [43] for completeness, then elaborate on our reshaping optimization.

Sparse Reshaping Model. The sparse reshaping model of human face stems from the forensic research of [11], which utilizes a linear regression model to present the mapping from the soft tissue thickness of 52 3D facial landmarks (Figure 3) to both the age and the body mass index (BMI). For a specific input portrait, the age remains unchanged. The positions of facial landmarks will change along their normal with respect to BMI differences (δBMI) as:

$$\begin{aligned} \boldsymbol{x}'_{I(j)} &= \boldsymbol{x}_{I(j)}(\boldsymbol{\alpha}_0) + \delta \mathrm{BMI} \cdot b_j \cdot \boldsymbol{n}_{I(j)}, \\ \boldsymbol{b} &= \{b_j\}, j = 1, 2, ..., 52, \end{aligned} \tag{4}$$

where $I(j)$ indicates the vertex index of the j-th facial landmark, $\boldsymbol{x}_{I(j)}(\boldsymbol{\alpha}_0), \boldsymbol{x}'_{I(j)}$ are the original and reshaped positions of the j-th facial landmark respectively, δBMI is the BMI difference, $\boldsymbol{n}_{I(j)} \in \mathbb{R}^3$ is the normal of the j-th facial landmark, and $\boldsymbol{b}$ consists of the regression coefficients of all facial landmarks.

Reshaping Optimization. The reshaping optimization can be formulated as the combination of deformation constraints encoding the sparse logical model in Eq. 4, and the constraints on face rationality and identity:

$$E(Y, \boldsymbol{\alpha}) = E_{shape}(Y) + E_{3DMM}(Y, \boldsymbol{\alpha}) + E_{ID}(Y), \tag{5}$$

where:

$$E_{shape}(Y) = \sum_j (\boldsymbol{y}_{I(j)} - \boldsymbol{x}'_{I(j)})^2, \tag{6}$$

is the sparse reshaping constraint. Then we define $E_{3DMM}(Y, \boldsymbol{\alpha})$ which ensures that the reshaped face lies in rational space:

$$E_{3DMM}(Y, \boldsymbol{\alpha}) = \sum_{i=1}^{n} w_i (\boldsymbol{x}_i(\boldsymbol{\alpha}) - \boldsymbol{y}_i)^2 + \frac{1}{2} \boldsymbol{\alpha}^T (\boldsymbol{\epsilon}^T \boldsymbol{\epsilon})^{-1} \boldsymbol{\alpha}, \tag{7}$$

where $\boldsymbol{\epsilon}$ is the diagonal matrix composed of eigenvalues of $\mathbf{P}_{id}$, the second term is the statistic regulation of identity coefficient $\boldsymbol{\alpha}$. Due to the representation limitation of 3DMM which will change face identity during optimization, we set a region Ω (red region in Figure 3a) as the region of interest for facial feature preservation, and assign a weight function as follows:

$$w_i = \begin{cases} 0.1, & if \boldsymbol{x}_i \in \Omega, \\ 1, & otherwise. \end{cases} \tag{8}$$

This is to set lower constraints of 3DMM to facial feature regions. Finally, the identity preservation term $E_{ID}(Y)$ which constrains the reshaped face to the original face, is defined as:

$$E_{ID}(Y) = \sum_{i=1}^{n} (\Delta(\boldsymbol{y}_i) - \Delta(\boldsymbol{x}_i(\boldsymbol{\alpha}_0)))^2, \tag{9}$$

and $\Delta(\cdot)$ is the vertex Laplacian operator of the 3D face mesh, which is simple and efficient enough to encode face identity since they are only tiny changes in facial features constrained by E_{shape}.

The formation of $E(Y, \boldsymbol{\alpha})$ allows a convex optimization, and the optimal solution is obtained solving a least-squares problem. Eventually, we obtain the optimal reshaped face Y^* respecting the body mass index δBMI, while preserving face identity and rationality. Finally, the facial expression encoded as $\boldsymbol{P}_{exp} \cdot \boldsymbol{\beta}_0$ is added to Y^* to obtain the resulting reshaped 3D face with expression.

6 SHAPEFACENET AND SHAPELY LEVEL ESTIMATOR

In addition to the reshaping algorithm, we design a shapely level estimator, which can automatically estimate how shapely a given portrait image is. Such an automatic estimator avoids tedious adjustments of reshaping parameters, allowing novice users to efficiently use our work. Thanks to the reshaping model we proposed in Section 5, we can simply utilize a single parameter δBMI to reasonably represent the change of a human face under weight variations. Moreover, the proposed reshaping model can synthesize different weight variations of a portrait image, which are hard to achieve. Hence we can generate a set of faces with different δBMI for a given portrait image. We ask human raters to select the most shapely or attractive face according to public aesthetic criteria, which is subjective and may form a shapely level distribution for the given portrait image. We chose the most voted one as the most shapely face. Then we can define the *shapely degree* of a face as the δBMI difference of the current face to the most shapely face. This naturally represents the shapely level of a portrait image.

Based on the above analysis, we constructed ShapeFaceNet, a dataset that consists of 3,400 individuals with shapely degree annotations. Each individual contains eight portrait images generated by our reshaping model and a shapely level distribution formed by the votes from 20 raters and a shapely degree (Figure 4) for each face, as discussed before. Please refer to Section B of supplementary material for more details.

We formulate the estimation of shapely degree as a regression problem. Inspired by the success of convolutional neural network in computer vision, we employ ResNet [17] to construct a regression model trained on ShapeFaceNet to estimate the shapely degree of a given portrait, which can then be used for reshaping portrait image.

7 WARPING OPTIMIZATION

In this step, we warp the input portrait image, such that the resultant 2D portrait is consistent with the 3D face model after reshaping, and is naturally blended with the background.

To this end, we propose a hybrid approach by warping the background (inspired by Shih et al. [35]) while seamlessly fusing with 3D reshaped face rendered to the image plane. The advantage is that both 2D image warping and 3D face rendering can be efficiently computed, which is much faster than the resolution-dependent warping method used in [43]. We describe the details of warping optimization in Section A of supplementary material, as this is not the main contribution of our method.

8 EVALUATION

In this section, we carefully evaluate the proposed method, including testing on various examples, presenting the performance and implementation details, comparing with the-state-of-the-art, conducting a pilot user study, and demonstrating its wide applications.

(a) $\lambda = -0.4$ (b) $\lambda = 0$ (c) $\lambda = 0.4$ (d) $\lambda = 0.8$ (e) $\lambda = 1.2$ (f) $\lambda = 1.6$ (g) $\lambda = 2.0$ (h) $\lambda = 2.4$

Figure 4: The annotated portraits of the same individual with increasing shapely degree after normalization.

Figure 5: Comparisons with the naive reshaping model. From left to right, input portraits, ours and the results of the naive reshaping model, respectively. The naive reshaping model generates noticeable warp artifacts around perspective sensitive regions such as eye region. ©Wei Xiang

8.1 Results

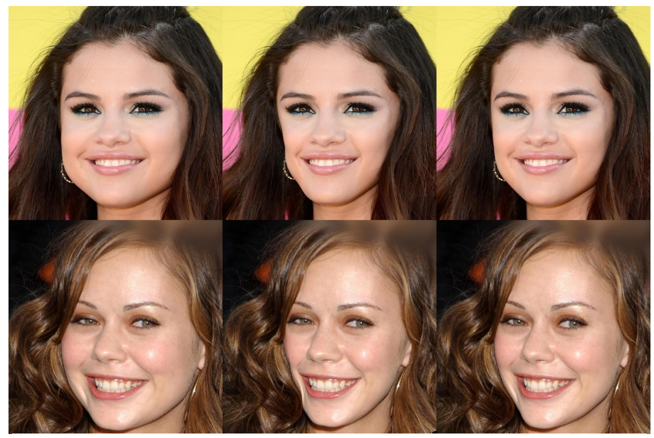
We have extensively tested the proposed method on portrait images with different poses (caused by varied camera angles) and expressions (calm, smile, grin, etc.) from FFHQ [21] and CelebA-HQ [20]. Figure 1 and Figure S4 in supplementary material exhibit various automatically generated results using test portrait images from the ShapeFaceNet dataset (see also the supplementary demo video). Note that for the last example in Figure 1, the man becomes more shapely toward larger facial roundness. It can be seen that despite the enormous variations (e.g., gender, hairstyle, pose, expression) of the input images, our method can generate satisfactory shapely portraits. The output portraits are highly realistic compared with the input images (see also the user study). Moreover, the result can be efficiently generated within a second, even for high-resolution inputs (see the next sub-section for detailed measurements).

8.2 Performance and Implementation Details

We have implemented our method on a desktop PC with Intel I7 4.0GHz CPU and 32GB memory. Our approach takes about 9ms for face reconstruction, 40ms for shapely level estimation, 6ms for reshaping, and 822ms for resolution-free warping optimization. Such performances demonstrate that our algorithm is suitable to be used in interactive applications.

8.3 Ablation Studies

For Eq. 5, the optimization seems more intuitive by setting $y_i = x_i(\alpha)$ to solve the 3DMM's parameter α only (noted as naive reshaping model), rather than using energy term E_{3DMM} (see Eq. 7) to directly solve the vertices of 3D face (our reshaping model). However, compared with the results of the naive reshaping model (see Figure 5), our reshaping model achieves more plausible results (especially on eye regions). Our reshaping model allows the reshaping face to be constrained by 3DMM, but not necessarily in the space of 3DMM, which has limited representations. Hence our reshaping model performs better.



(a) Input (b) Zhao et al. [43] (c) Our method

Figure 6: For the same inputs (a), we evaluate our method against [43] with the same (high) reshaping parameter. Prior work generates irrational reshaped portraits with wide forehead, narrow chin and cheek, and severe artifacts around face contour (b), while our method can still generate natural and plausible results (c).

8.4 Comparisons to the State-of-the-Art

We compare our deep shapely portrait approach against the state-of-the-art reshaping method of Zhao et al. [43]. Since the user must manually specify the adjustment parameter in their method, we compare the results under the same reshaping parameter according to the shapely degree automatically estimated by our predictor.

We design three sets for comparison to validate the effectiveness of our reshaping model. The first comparison is on frontal view portrait images. As shown in the first row of Figure 6, their method generates artifacts typically with a broad forehead, a narrow chin, and a narrow cheek. In contrast, our approach generates harmonious and natural faces. The second experiment is on profile portrait images. As shown in the second row of Figure 6, for large adjustments, their method generates noticeable artifacts around facial contours, while our results keep natural face boundary. Figure S3 in supplementary material further illustrates the reshaping

Figure 7: Comparisons with Meitu and Facetune2. From top to bottom, input portraits, our, Meitu, and Facetune2 results, respectively. The artifacts of Facetune2 are highlighted in green box.

faces under multiple shapely degrees in increasing order. Comparisons show that the method of Zhao et al. [43] generates irrational changes at the chin and cheek regions when the shapely degrees are negative. In contrast, our approach can always make natural-looking faces. In general, our method exhibits superior performance. This is mainly due to the involvement of rational face space constraints based on 3DMM and identity constraints in the reshaping model, not considered by prior work.

Figure S1 in supplementary material shows the comparison of computational time. For an image with 4 megapixels (MP), our approach is about 10 times faster. Moreover, our optimization time (about 0.8 seconds) is resolution-free, indicating that our method can efficiently process portraits with very high resolution. For portrait images with more than 4 MP, their method simply fails because of memory overflow.

8.5 User Study

We have conducted four user studies to evaluate our result by comparing: (a) our result to the original portrait (noted as **Our-OG**), (b) our result to the annotated shapely portrait in ShapeFaceNet (noted as **Our-AS**), (c) our result to the reshaping result of two popular portrait and selfie editor Apps, Meitu [31] and Facetune2 [12] (noted as **Our-Apps**). Moreover, we verify whether a human subject and his/her friends like his/her shapely portraits (generated by our method) compared with the original one (noted as **SF**).

For the first two studies, we randomly selected 50 individuals from test images of the ShapeFaceNet. We recruited 34 subjects for **Our-OG** and 31 subjects for **Our-AS**. In each test, we asked the subject to view 50 pairs of images of the 50 individuals mentioned above, and then select the more shapely one in each image pair. For **Our-OG**, each image pair is formed by the original image and the automatic reshaping result of the same individual. For **Our-AS**, each pair consists of the automatic reshaping result and the annotated shapely portrait of the same individual. In each pair, there is no specific order for the two images. As summarized in Table 1, for **Our-OG**, 74% of the subjects chose our result, which proves the effectiveness of our method in generating shapely portrait. For **Our-AS**, 44.6% and 55.4% chose our result and the annotated shaped portrait (can be treated as ground-truth), respectively, indicating that our deep-learning-based approach is capable of reshaping the given portrait according to public aesthetic criteria.

For the third study, we selected Meitu and Facetune2, two popular commercial tools on iOS & Android for portrait and selfie editing, according to the ranking list of the most downloaded Apps. Figure 7 shows six input portraits (first row) and the results generated by our method (second row), Meitu (third row), and Facetune2 (last row). Although all results are based on face reshaping, our result is generated automatically while the results of Meitu and Facetune2 are generated with the help of artists by manually adjusting input portraits to be shapely. It can be seen that our method allows for more massive face deformations that still look natural and plausible. To better evaluate the quality of the results according to human aesthetics, 24 triples of portraits were randomly presented to 37 raters who were asked to rank the quality according to the shapeliness of the presented portrait. Statistics show that 59.12%, 23.65%, 17.23% raters chose our result, Meitu, Facetune2 as the most shapely one, respectively. This study validates that our approach can generate more shapely results. Moreover, we found that Facetune2 can easily lead to noticeable artifacts around facial contours and hair regions, which are highlighted in green boxes.

For the last study (**SF**), we verify whether a human subject and his/her friends like his/her shapely portraits (generated by our method). We recruit 24 subjects (4 males and 20 females aging from 19 to 30) for this task. Each subject and 5 of his/her friends view the original and shapely portraits and select the better-looking one between the two. As shown in Table 1, 16.7% (21.7%) and 83.3% (78.3%) of the subjects (their friends) prefer the original portrait and our shapely portrait, respectively, indicating that our method can generate more shapely portraits even for observers that have a high familiarity to the test subjects.

Table 1: User study results for our automatic reshaping model with three settings (Our-OG, Our-AS, and SF).

Number of samples	Preference	
Our-OG	Original portrait	Our result
1,700	442 (26%)	1,258 (74%)
Our-AS	Annotated shaped portrait	Our result
1,550	859 (55.4%)	691 (44.6%)
SF	Original portrait	Our result
Self 24	4 (16.7%)	20 (83.3%)
Friends 120	26 (21.7%)	94 (78.3%)

(a) Input (b) Retouching (c) Retouching + Reshaping

Figure 8: Given an input portrait image (a), texture filtering only adjusts face colors (b), while our method further enhances the portrait by face reshaping (c).

(a) (b) (c)

Figure 9: Our method cannot improve double chin and extra fat at neck for the examples shown in (a) and (b) as our method only changes the face shape. For some examples as in (c), our shapely portrait may not look that attractive as our approach does not refine face texture.

8.6 Applications

Our method enables automatic/manual adjustment of a human face, producing satisfactory shapely portrait images. It can be easily incorporated with various portrait editing methods to make the portrait look gorgeous. Therefore, it has a wide range of applications on social media, digital entertainment, film/television production, and facelift. Social Apps (such as Facebook, Instagram, etc.) also provide tools to retouch photos using filters. As our method can naturally reshape portraits, it can be integrated seamlessly therein. Figure 8 demonstrates three examples which combine our shapely editing with texture filtering (we use the builtin retouching tools in Instagram). Given a portrait image as shown in Figure 8(a), we first retouch it using the texture filtering tool in Instagram (see Figure 8(b)), then reshape it using our method (see Figure 8(c)). The results show that our shapely editing method can significantly increase the attractiveness of the results generated by filtering only.

9 DISCUSSION

While our automated method can achieve shapely portraits with high quality and realism, it still has some restrictions. First, we only reconstruct the face regions with a simple parametric model, which does not include other body parts in the portrait images. Thus we can only reshape face but not neck and chin, for example. For obese people that usually have a double chin and extra fat at their neck (see Figure 9(a)(b)), it would be better also to consider these regions. Second, we do not adjust the lighting and texture of the input portrait. The result may not be that impressive, as shown in Figure 9(c). This can be overcome by using image editing tools such as Adobe Photoshop and relighting techniques such as [38], which is beyond the scope of the current work. Third, input portraits with invisible camera distortion are essential for our method to generate plausible shapely results. Like all other image retargeting methods, our warping optimization may generate subtle artifacts at the background under notable reshaping adjustment. And it is currently the efficiency bottleneck of our approach by taking over 90% of computational time. Further, how to quantitatively evaluate the resultant shapely portrait using a typical image error metric (e.g., RMSE) remains a problem.

10 CONCLUSION AND FUTURE WORK

In this paper, we present deep shapely portrait, the first automatic, learning-based method to reshape a given portrait to be better proportioned and more shapely. The proposed method is based on a novel CNN-based shapely level estimator and uses the estimated shapely degree to reshape face in 3D. To achieve natural face deformation, the reshaping model leverages the underlying shape component of the reconstructed face based on 3D morphable models. The soft tissue thickness regression model from a specific set of facial landmarks is collected from human faces. Finally, a resolution-free image warping is proposed to blend the reshaped face with the background seamlessly. Through extensive experiments and user studies, we have shown that our method outperforms previous work and achieves stunning results. We also demonstrate two applications by combing our method with other portrait editing work and commercial software. We see our work as a significant step towards natural and realistic portrait image editing and hope it can inspire other works in the field.

In the future, we would like to make the shapely level estimation more general by augmenting the training set to handle more variations (e.g., skin color), and further extend our method to generate shapely portrait videos. How to enable less supervision by directly/consistently labeling portrait images of different people still remains a challenge. Also, complementing the current face reshaping model by including other parametric models such as neck [30] is an interesting direction to explore.

11 ACKNOWLEDGMENTS

We thank Wei Xiang and Hanqi Lv for photo usage permissions. Xiaogang Jin was supported by the National Key R&D Program of China (Grant No. 2017YFB1002600), the National Natural Science Foundation of China (Grant No. 61972344), and the Key Research and Development Program of Zhejiang Province (Grant No. 2018C03055). Yong-Liang Yang was supported by CAMERA - the RCUK Centre for the Analysis of Motion, Entertainment Research and Applications (EP/M023281/1), and a gift from Adobe.

REFERENCES

[1] Jascha Achenbach, Eduard Zell, and Mario Botsch. 2015. Accurate Face Reconstruction through Anisotropic Fitting and Eye Correction. In *Vision, Modeling & Visualization*, David Bommes, Tobias Ritschel, and Thomas Schultz (Eds.). The Eurographics Association. https://doi.org/10.2312/vmv.20151251

[2] Volker Blanz and Thomas Vetter. 1999. A Morphable Model for the Synthesis of 3D Faces. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, Vol. 99. 187–194. https://doi.org/10.1145/311535.311556

[3] James Booth, Anastasios Roussos, Allan Ponniah, David Dunaway, and Stefanos Zafeiriou. 2018. Large Scale 3D Morphable Models. *International Journal of Computer Vision* 126, 2-4 (2018), 233–254. https://doi.org/10.1007/s11263-017-1009-7

[4] Chen Cao, Derek Bradley, Kun Zhou, and Thabo Beeler. 2015. Real-Time High-Fidelity Facial Performance Capture. *ACM Transactions on Graphics* 34, 4, Article 46 (July 2015), 9 pages. https://doi.org/10.1145/2766943

[5] Chen Cao, Yanlin Weng, Shun Zhou, Yiying Tong, and Kun Zhou. 2014. FaceWarehouse: A 3D Facial Expression Database for Visual Computing. *IEEE Transactions on Visualization and Computer Graphics* 20, 3 (March 2014), 413–425. https://doi.org/10.1109/TVCG.2013.249

[6] Menglei Chai, Linjie Luo, Kalyan Sunkavalli, Nathan Carr, Sunil Hadap, and Kun Zhou. 2015. High-Quality Hair Modeling from a Single Portrait Photo. *ACM Transactions on Graphics* 34, 6, Article 204 (Oct. 2015), 10 pages. https://doi.org/10.1145/2816795.2818112

[7] Menglei Chai, Tianjia Shao, Hongzhi Wu, Yanlin Weng, and Kun Zhou. 2016. AutoHair: Fully Automatic Hair Modeling from a Single Image. *ACM Transactions on Graphics* 35, 4, Article 116 (July 2016), 12 pages. https://doi.org/10.1145/2897824.2925961

[8] Che-Han Chang, Yoichi Sato, and Yung-Yu Chuang. 2014. Shape-Preserving Half-Projective Warps for Image Stitching. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 3254–3261.

[9] Fangmei Chen, Xihua Xiao, and David Zhang. 2018. Data-Driven Facial Beauty Analysis: Prediction, Retrieval and Manipulation. *IEEE Transactions on Affective Computing* 9, 2 (2018), 205–216. https://doi.org/10.1109/TAFFC.2016.2599534

[10] Yu-Sheng Chen and Yung-Yu Chuang. 2016. Natural Image Stitching with the Global Similarity Prior. In *European Conference on Computer Vision*. Springer, 186–201. https://doi.org/10.1007/978-3-319-46454-1_12

[11] Sven De Greef, Peter Claes, Dirk Vandermeulen, Wouter Mollemans, Paul Suetens, and Guy Willems. 2006. Large-Scale In-Vivo Caucasian Facial Soft Tissue Thickness Database for Craniofacial Reconstruction. *Forensic science international* 159 (2006), S126–S146. https://doi.org/10.1016/j.forsciint.2006.02.034

[12] Facetune2. 2020. Facetune2 - Best Seflie Editor App. www.facetuneapp.com.

[13] Yangyu Fan, Shu Liu, Bo Li, Zhe Guo, Ashok Samal, Jun Wan, and Stan Z. Li. 2018. Label Distribution-Based Facial Attractiveness Computation by Deep Residual Learning. *IEEE Transactions on Multimedia* 20, 8 (2018), 2196–2208. https://doi.org/10.1109/TMM.2017.2780762

[14] Junying Gan, Lichen Li, Yikui Zhai, and Yinhua Liu. 2014. Deep Self-Taught Learning for Facial Beauty Prediction. *Neurocomputing* 144 (2014), 295–303. https://doi.org/10.1016/j.neucom.2014.05.028

[15] Lian Gao, Weixin Li, Zehua Huang, and Di Huang. 2018. Automatic Facial Attractiveness Prediction by Deep Multi-Task Learning. In *International Conference on Pattern Recognition (ICPR)*. 3592–3597. https://doi.org/10.1109/ICPR.2018.8545033

[16] Xiaoguang Han, Kangcheng Hou, Dong Du, Yuda Qiu, Shuguang Cui, Kun Zhou, and Yizhou Yu. 2018. CaricatureShop: Personalized and Photorealistic Caricature Sketching. *IEEE Transactions on Visualization and Computer Graphics* (2018), 1–1. https://doi.org/10.1109/TVCG.2018.2886007

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 770–778.

[18] Liwen Hu, Shunsuke Saito, Lingyu Wei, Koki Nagano, Jaewoo Seo, Jens Fursund, Iman Sadeghi, Carrie Sun, Yen-Chun Chen, and Hao Li. 2017. Avatar Digitization from a Single Image for Real-Time Rendering. *ACM Transactions on Graphics* 36, 6, Article 195 (Nov. 2017), 14 pages. https://doi.org/10.1145/3130800.31310887

[19] Patrik Huber, Guosheng Hu, Rafael Tena, Pouria Mortazavian, P Koppen, William J Christmas, Matthias Ratsch, and Josef Kittler. 2016. A Multiresolution 3D Morphable Face Model and Fitting Framework. In *Proceedings of the 11th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. INSTICC, 79–86. https://doi.org/10.5220/0005669500790086

[20] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. 2018. Progressive Growing of GANs for Improved Quality, Stability, and Variation. In *International Conference on Learning Representations*.

[21] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8110–8119.

[22] Vahid Kazemi and Josephine Sullivan. 2014. One Millisecond Face Alignment with an Ensemble of Regression Trees. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1867–1874.

[23] Hyeongwoo Kim, Pablo Carrido, Ayush Tewari, Weipeng Xu, Justus Thies, Matthias Niessner, Patrick Pérez, Christian Richardt, Michael Zollhöfer, and Christian Theobalt. 2018. Deep Video Portraits. *ACM Transactions on Graphics* 37, 4, Article 163 (2018), 14 pages. https://doi.org/10.1145/3197517.3201283

[24] Tommer Leyvand, Daniel Cohen-Or, Gideon Dror, and Dani Lischinski. 2008. Data-Driven Enhancement of Facial Attractiveness. *ACM Transactions on Graphics* 27, 3, Article 38 (Aug. 2008), 9 pages. https://doi.org/10.1145/1360612.1360637

[25] Honglin Li, Masahiro Toyoura, and Xiaoyang Mao. 2019. Caricature Synthesis with Feature Deviation Matching under Example-Based Framework. *The Visual Computer* 35, 5 (2019), 653–666. https://doi.org/10.1007/s00371-018-1495-9

[26] Tianye Li, Timo Bolkart, Michael J. Black, Hao Li, and Javier Romero. 2017. Learning a Model of Facial Shape and Expression from 4D Scans. *ACM Transactions on Graphics* 36, 6, Article 194 (Nov. 2017), 17 pages. https://doi.org/10.1145/3130800.3130813

[27] Qiqi Liao, Xiaogang Jin, and Wenting Zeng. 2012. Enhancing the Symmetry and Proportion of 3D Face Geometry. *IEEE Transactions on Visualization and Computer Graphics* 18, 10 (2012), 1704–1716. https://doi.org/10.1109/TVCG.2012.26

[28] Si Liu, Xinyu Ou, Ruihe Qian, Wei Wang, and Xiaochun Cao. 2016. Makeup Like a Superstar: Deep Localized Makeup Transfer Network. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI'16)*. AAAI Press, 2568–2575.

[29] Xudong Liu, Tao Li, Hao Peng, Iris Chuoying Ouyang, Taehwan Kim, and Ruizhe Wang. 2019. Understanding Beauty via Deep Facial Features. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[30] Yilong Liu, Chengwei Zheng, Feng Xu, Xin Tong, and Baining Guo. 2020. Data-Driven 3D Neck Modeling and Animation. *IEEE Transactions on Visualization and Computer Graphics* (2020). https://doi.org/10.1109/TVCG.2020.2967036

[31] Meitu. 2020. Meitu – Beauty Cam, Easy Photo Editor. https://play.google.com/store/apps/details?id=com.mt.mtxx.mtxx&hl=en.

[32] Elad Richardson, Matan Sela, Roy Or-El, and Ron Kimmel. 2017. Learning Detailed Face Reconstruction from a Single Image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1259–1268.

[33] Rasmus Rothe, Radu Timofte, and Luc Van Gool. 2016. Some Like It Hot - Visual Guidance for Preference Prediction. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 5553–5561.

[34] Omry Sendik, Dani Lischinski, and Daniel Cohen-Or. 2019. What's in a Face? Metric Learning for Face Characterization. In *Computer Graphics Forum*, Vol. 38. Wiley Online Library, 405–416. https://doi.org/10.1111/cgf.13647

[35] YiChang Shih, Wei-Sheng Lai, and Chia-Kai Liang. 2019. Distortion-Free Wide-Angle Portraits on Camera Phones. *ACM Transactions on Graphics* 38, 4, Article 61 (2019), 12 pages. https://doi.org/10.1145/3306346.3322948

[36] YiChang Shih, Sylvain Paris, Connelly Barnes, William T. Freeman, and Frédo Durand. 2014. Style Transfer for Headshot Portraits. *ACM Transactions on Graphics* 33, 4, Article 148 (July 2014), 14 pages. https://doi.org/10.1145/2601097.2601137

[37] Zhixin Shu, Sunil Hadap, Eli Shechtman, Kalyan Sunkavalli, Sylvain Paris, and Dimitris Samaras. 2017. Portrait Lighting Transfer Using a Mass Transport Approach. *ACM Transactions on Graphics* 37, 1, Article 2 (Oct. 2017), 15 pages. https://doi.org/10.1145/3095816

[38] Tiancheng Sun, Jonathan T. Barron, Yun-Ta Tsai, Zexiang Xu, Xueming Yu, Graham Fyffe, Christoph Rhemann, Jay Busch, Paul Debevec, and Ravi Ramamoorthi. 2019. Single Image Portrait Relighting. *ACM Transactions on Graphics* 38, 4, Article 79 (July 2019), 12 pages. https://doi.org/10.1145/3306346.3323008

[39] Ayush Tewari, Florian Bernard, Pablo Garrido, Gaurav Bharaj, Mohamed Elgharib, Hans-Peter Seidel, Patrick Pérez, Michael Zollhofer, and Christian Theobalt. 2019. FML: Face Model Learning from Videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 10812–10822.

[40] Anh Tuan Tran, Tal Hassner, Iacopo Masi, and Gérard Medioni. 2017. Regressing Robust and Discriminative 3D Morphable Models with a very Deep Neural Network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5163–5172.

[41] Duorui Xie, Lingyu Liang, Lianwen Jin, Jie Xu, and Mengru Li. 2015. SCUT-FBP: A Benchmark Dataset for Facial Beauty Perception. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*. 1821–1826. https://doi.org/10.1109/SMC.2015.319

[42] Jie Xu, Lianwen Jin, Lingyu Liang, Ziyong Feng, Duorui Xie, and Huiyun Mao. 2017. Facial Attractiveness Prediction Using Psychologically Inspired Convolutional Neural Network (PI-CNN). In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 1657–1661. https://doi.org/10.1109/ICASSP.2017.7952438

[43] Haiming Zhao, Xiaogang Jin, Xiaojian Huang, Menglei Chai, and Kun Zhou. 2018. Parametric Reshaping of Portrait Images for Weight-Change. *IEEE Computer Graphics and Applications* 38, 1 (2018), 77–90. https://doi.org/10.1109/MCG.2018.011461529

[44] Shizhe Zhou, Hongbo Fu, Ligang Liu, Daniel Cohen-Or, and Xiaoguang Han. 2010. Parametric Reshaping of Human Bodies in Images. *ACM Transactions on Graphics* 29, 4, Article 126 (2010), 10 pages. https://doi.org/10.1145/1778765.1778863